September 1966

LOWER BOUNDS FOR AVERAGE SAMPLE NUMBER OF SEQUENTIAL MULTIHYPOTHESIS TESTS

Gordon Simons

Technical Report No. 81

University of Minnesota

Minneapolis, Minnesota

*Research supported under a grant from the United States Air Force Office of Scientific Research - Office of Aerospace Research, Grant #AF-AFOSR-885-65.

LOWER BOUNDS FOR AVERAGE SAMPLE NUMBER

OF SEQUENTIAL MULTIHYPOTHESIS TESTS

By

Gordon Simons

Summary.

Sections 1-5 are concerned with finding lower bounds for the expected sample sizes of sequential multihypothesis tests in the presence of a constraining error matrix. We consider K simple hypotheses corresponding to K density functions f_i, $i=1,\ldots,K$, and fix all of the entries of the $K \times K$ error matrix $A = (\alpha_{ij})$, where α_{ij} = P[accepting $f_j | f_i$ true]. Lower bounds are found for $E(N|f)$, first, when f is one of the K densities, and then, for a $K+1^{st}$ density f_0. In section 6, lower bounds are found when the error constraints arising from the error matrix are relaxed and/or modified. Section 7 finds lower bounds for average sample size when the test is not constrained by an error matrix but rather by a lower bound for the probability of a "correct decision" as a function of the true state of nature.

1. Introduction.

Let $X_1, X_2, \ldots$ be a sequence of independent random variables having a common density function f with respect to some σ-finite measure μ. Consider a test of hypotheses where H_ν is the hypothesis that $f=f_\nu$, $\nu=1,\ldots,K$, with $K \geq 2$. Let α_{ij} = P[accepting $H_j | H_i$]. When it is needed in the discussion, we will let f_0 be a $K+1^{st}$ density with respect to μ. Let N denote the (random) number of observations taken by the test. We are concerned with finding lower bounds for $E_\nu(N)$ subject to the constraining error matrix $A = (\alpha_{ij})$ under density f_ν, $\nu=1,\ldots,K$ or $\nu = 0$.

The history of this problem is as follows. Frequently f is a density depending on some parameter θ, so that f_ν corresponds to some density with parameter θ_ν, $\nu=0,1,\ldots,K$. When $K = 2$, Wald's sequential probability ratio test (SPRT) for testing θ_1 against θ_2 with error probabilities $\alpha_{12} = \alpha$ and $\alpha_{21} = \beta$, minimizes $E_1(N)$ and $E_2(N)$ (A. Wald and J. Wolfowitz [17]). However, in practice, the true parameter may be a third value θ_0 and $E_0(N)$ might be quite large. For instance, if f is the normal density with mean θ, choosing $\theta_0 = (\theta_1+\theta_2)/2$ and sufficiently small α and β, we can make $E_0(N)$ for the SPRT exceed the sample size in the usual fixed sample size test. This unpleasant situation has encouraged the development of other sequential tests such as a "modified SPRT" given by T.W. Anderson [1]. The goal for such a test is to keep $E(N|\theta)$ near the ASN of the SPRT when $\theta = \theta_1$ or θ_2 and as small as possible for other θ's or at least for the most "objectionable" θ's.

Clearly, a useful criterion for evaluating the performance of a sequential test is to compare the ASN with a good theoretical lower bound for ASN. A favorable comparison enhances both the test and the lower bound while an unfavorable one is not conclusive. Anderson showed that his modified SPRT could produce a favorable comparison.

The problem of developing theoretical lower bounds for ASN has been treated rather satisfactorily for $K = 2$ by A. Wald [14] and W. Hoeffding ([7], [8]). Wald showed that

$$E_1(N) \geq \frac{(1-\alpha)\ln((1-\alpha)/\beta) + \alpha\ln(\alpha/(1-\beta))}{\int f_1 \ln(f_1/f_2)d\mu} \tag{1.1}$$

and

$$E_2(N) \geq \frac{\beta\ln(\beta/(1-\alpha)) + (1-\beta)\ln((1-\beta)/\alpha)}{\int f_2 \ln(f_2/f_1)d\mu} . \tag{1.2}$$

Wald's proof is given for nonrandomized tests, but it extends to randomized tests as well (See lemmas 1 and 3 in Section 2).

When neither hypothesis is true, Hoeffding has given three different lower bounds for ASN under f_0. In his 1953 paper [7], he gave the lower bound

$$E_0(N) \geq \sup_{0<d<1} \frac{-\ln[(1-\alpha)^d\beta^{(1-d)} + \alpha^d(1-\beta)^{(1-d)}]}{d\int f_0 \ln(f_0/f_1)d\mu + (1-d)\int f_0 \ln(f_0/f_2)d\mu} . \tag{1.3}$$

In his 1960 paper [8], he gave two lower bounds

$$E_0(N) \geq \frac{1 - \alpha - \beta}{1-\int \min(f_0, f_1, f_2)d\mu} , \tag{1.4}$$

and

$$E_0(N) \geq \frac{\{[(\tau/4)^2 - \zeta\ln(\alpha+\beta)]^{\frac{1}{2}} - (\tau/4)\}^2}{\zeta^2} , \tag{1.5}$$

where

$$(1.6)\qquad \zeta = \max(\zeta_1,\zeta_2),\ \zeta_i = \int f_0 \ln(f_0/f_i)d\mu, \quad i=1,2,$$

and

$$(1.7)\qquad \tau^2 = \int(\ln(f_2/f_1) - \zeta_1+\zeta_2)^2 f_0 d\mu \ .$$

Sequential tests with three or more hypotheses may be evaluated by generalizing the previous results to K hypotheses. Wald's lower bounds, (1.1) and (1.2), extend for K hypotheses to the important bound

$$(1.8)\qquad E_i(N) \geq \max_{\substack{1\leq j\leq K\\ j\neq i}} \frac{\sum_{\nu=1}^{K} \alpha_{i\nu}\ln(\alpha_{i\nu}/\alpha_{j\nu})}{\int f_i \ln(f_i/f_j)d\mu} \quad , \quad i=1,\ldots,K \ .$$

This bound leads to a disguised generalization of Hoeffding's bound (1.3), namely,

$$(1.9)\qquad E_0(N) \geq \inf_{\{b_\nu\}} \max_{1\leq j\leq K} \frac{\sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_j)d\mu} \ ,$$

where

$$(1.10)\qquad \sum_{1}^{K} b_\nu = 1;\ b_\nu > 0, \quad \text{for}\quad \nu=1,\ldots,K \ .$$

Hoeffding's bound (1.5) extends to

$$(1.11)\qquad E_0(N) \geq [(T^2 - R\ln S)^{\frac{1}{2}} - T]^2/R^2$$

where R, S, and T are defined for subsets (size two or larger) of the first K positive integers. Let D be such a subset with ν members; let $C = \{C_i \mid i\epsilon D\}$ be a set of ν real numbers for which

$$(1.12) \qquad \sum_{i\in D} C_i = 0 \ , \ \sum_{i\in D} |C_i| = 1 \ ;$$

and let $\Phi(D)$ be the permutations of D with typical member φ (a ν-dimentional vector). Then

$$(1.13) \qquad R(D) = \max_{i\in D} \int f_0 \ln(f_0/f_i) d\mu \ ,$$

$$(1.14) \qquad S(D) = \sum_{j=1}^{K} \min_{i\in D} \alpha_{ij} \ ,$$

and

$$(1.15) \qquad T(D) = \inf_{C} \tau(C)/2\nu(\nu-2)! \ ,$$

where

$$\tau(C) = \sum_{\varphi\in\Phi(D)} \tau_\varphi(C) \ ,$$

with

$$\tau_\varphi^2(C) = \int f_0 \Big(\sum_{i\in D} C_{\varphi_i} [\ln(f_0/f_i) - E_0 \ln(f_0/f_i)] \Big)^2 d\mu \ .$$

Bound (1.11) is really several bounds when $K > 2$, one for each subset D. To derive the bound we must make a regularity assumption. This will be given in section 5 when the bound is verified. Substantial improvement on Hoeffding's conditions may be noted.

Presumably, there is a generalization of bound (1.4), but it is doubtful that any generalization would be of much value in typical applications. For instance, if f_i is the density coming from a normal distribution $N(\theta_i,\sigma^2)$ $(i=0,1,2)$ and $\theta_1 < \theta_2$, then bound (1.4) is a constant for all θ_0 in the interval $[\theta_1,\theta_2]$ while the graphs of bounds

(1.3) and (1.5) bulge upward substantually over the same interval. Another serious objection to bound (1.4) arises as follows. Let f_0, f_1, f_2 be normal densities with means 0, $-\delta$, and δ respectively and common variance. Then bound (1.4) is of order δ^{-1} while the other bounds ((1.3) and (1.5)) are proportional to δ^{-2} and hence, are better for small δ. (The ASN at $\theta = 0$ for the SPRT when testing $\theta = -\delta$ against $\theta = \delta$ with fixed α and β is also proportional to δ^{-2}.)

Section 2 presents lemmas which are used in subsequent sections. Sections 3, 4, and 5 verify and discuss bounds (1.8), (1.9), and (1.11) respectively.

Bounds (1.8), (1.9), and (1.11) are predicated on complete control of an error matrix. For given error matrix, a sequential test usually can be found by introducing extensive randomization. (See Theorem 4.2 of Section 4.) This being objectionable, typical applications are based on error matrices which are only partially controlled. Section 6 discusses this problem and finds some additional lower bounds for ASN. A table compares one of these lower bounds with the actual ASN of a three hypothesis test which the author [11] has investigated.

Authors such as M. Sobel and A. Wald [12] as well as E. Paulson [9] have adopted a "correct decision" approach to multihypothesis testing rather than use the error matrix approach. Briefly, the parameter space is initially partitioned into K disjoint sets $S_1,\ldots,S_K$ corresponding to K hypotheses $H_1,\ldots,H_K$. Then "indifference regions" are introduced which have the effect of increasing the sets to new overlapping ones $S_1',\ldots,S_K'$. A test terminates in acceptance of one of the hypotheses H_j.

One makes a "correct decision" if S'_j contains the true state θ. Finally, one insists that the probability of a correct decision must be as large as some value P^* for all θ in the parameter space. P^* may depend on θ. Section 7 finds lower bounds for ASN in this case and makes some numerical comparisons with the three hypothesis test given by Sobel and Wald [12].

2. Lemmas for subsequent sections.

Lemmas 1, 2, and 3 are lemmas involving stopping variables. By the term stopping variable, with respect to a sequence of random variables $X_1, X_2, \cdots,$ we will mean a random variable N defined on the non-negative integers such that the occurrence of the event $[N=n]$ depends at most on $X_1, \cdots, X_n$ and on a randomization based on the values of $X_1, \cdots, X_n$ for $n=1,2,\cdots,$ and on none of the X's for $n=0$.

Lemma 1.[1/] *Let* $X_1, X_2, \cdots,$ *be a sequence of independent random variables identically distributed as* Z *and let* N *be a stopping variable (with respect to the* X's). *If* $E(N) < \infty$ *and* $E(|X|) < \infty$, *then*

$$E(Z_N) = E(N)E(Z) , \tag{2.1}$$

where

$$Z_n = \sum_1^n X_i . \tag{2.2}$$

1/ (2.1) has been proven under various conditions and in various ways. e.g. Wald ([13],[14]), Blackwell [2], Blackwell and Girshick [3], and Doob [6], pp. 350-351. None of these considered N as randomized.

Lemma 2. [1/] Let $X_1, X_2, \ldots,$ be a sequence of independent random variables identically distributed as Z and let N be a stopping variable (with respect to the X's). Let $E(Z) = 0$. If $E(N) < \infty$ and $E(Z^2) < \infty$, then

$$E(Z_N^2) = E(N)E(Z^2) \tag{2.3}$$

where

$$Z_n = \sum_1^n X_i \ . \tag{2.4}$$

We can modify proofs by Doob [6] and Chow-Robbins-Teicher [5] to prove lemmas 1 and 2 respectively. Both proofs use martingales which can easily be modified to include randomization.

Lemma 3. Let $X_1, X_2, \ldots$ be a sequence of random variables. Let H_i be the hypothesis that $X_1, \ldots, X_n$ have joint density f_{in} with respect to some measure μ_n for $n=1,2,\ldots,$ and for $i=1,\ldots,K$. Let N be the total number of observations (a stopping variable) in any sequential closed (under all hypotheses) sampling scheme. Let $E_i^\nu(\cdot)$ denote expectation under H_i conditional on terminal acceptance of H_ν for $i,\nu=1,\ldots,K$. Assume $\alpha_{i\nu} > 0$. Then

$$E_i^\nu\left(\frac{f_{jN}}{f_{iN}}\right) \le \frac{\alpha_{j\nu}}{\alpha_{i\nu}} \quad \text{for} \quad i,j,\nu=1,\ldots,K(\alpha_{k\ell} \equiv P_k[\text{accepting } H_\ell]) \ . \tag{2.5}$$

Equality holds if $f_{jn} = 0$ whenever $f_{in} = 0$. When $N = 0$, $\frac{f_{jN}}{f_{iN}}$ is defined as unity.

1/ (2.3) has been proven under various conditions and in various ways. See e.g. Wald ([15],[16]), Wolfowitz [18], and Chow-Robbins-Teicher [5]. None of these considered N as randomized.

Proof. We can represent the above testing scheme by the pair (ψ,ϕ) $\psi = (\psi_0,\psi_1,\ldots)$ is an infinite dimensional vector with $\psi_n = \psi_n(X) = P[N=n|X]$ where $X = (X_1,X_2,\ldots)$ and where the dependence on X is through $X_1,\ldots,X_n$ only for $n = 0,1,\ldots$. $\phi = (\phi_{ni})$ is an infinite by finite dimensional matrix with

$$\phi_{ni} = \phi_{ni}(X) = P[\text{accepting } H_i | X \text{ and } N = n]$$

where the dependence on X is again through $X_1,\ldots,X_n$ only for $n=0,1,\ldots,$ and $i=1,\ldots,K$. Let A_n be the event $[f_{in} \neq 0]$ for $n \geq 1$. Then

$$E_i^\nu\left(\frac{f_{jN}}{f_{iN}}\right) = \frac{\psi_0\phi_{0\nu}\cdot(1) + \sum_{n=1}^{\infty} \int_{A_n} \psi_n\phi_{n\nu} \cdot \frac{f_{jn}}{f_{in}} \cdot f_{in}d\mu_n}{P_i[\text{accepting } H_\nu]}$$

$$\leq \frac{\psi_0\phi_{0\nu} + \sum_{n=1}^{\infty} \int \psi_n\phi_{n\nu}f_{jn}d\mu_n}{\alpha_{i\nu}}$$

$$= \frac{\alpha_{j\nu}}{\alpha_{i\nu}} \quad .$$

Lemmas 4, 5, and 6 are lemmas related to information theory. Specifically, lemmas 4 and 5 are proven by C. R. Rao [10], pg. 47 and lemma 6 is a corollary of lemma 5.

Lemma 4. *Let f and g be two density functions with respect to the same measure μ. If $f = 0$ whenever $g = 0$ then*

(2.6) $$\int f \ln(f/g)d\mu \geq 0$$

with equality holding if, and only if, $f = g$ a.e. (μ). ($0 \ln(0/g)$ *is to be interpreted as zero.*)

Lemma 5. *Let* $a_1,\ldots,a_n$ *and* $b_1,\ldots,b_n$ *be two sequences of positive real numbers for which*

$$(2.7) \qquad \sum_1^n a_i = \sum_1^n b_i = 1 .$$

Then

$$(2.8) \qquad \sum_1^n a_i \ln(a_i/b_i) \geq 0 ,$$

with equality holding if, and only if $a_i = b_i$ *for all* i.

Lemma 6. *Let* $a_1,\ldots,a_n$ *and* $b_1,\ldots,b_n$ *be two sequences of positive real numbers.* *Let*

$$(2.9) \qquad a = \sum_1^n a_i \quad \text{and} \quad b = \sum_1^n b_i .$$

Then

$$(2.10) \qquad \sum_1^n a_i \ln a_i/b_i \geq a \ln a/b$$

with equality holding if, and only if, $\frac{a_i}{b_i} = \frac{a}{b}$ *for all* i.

Proof. Normalizing, the two sequences become $a_1/a,\ldots,a_n/a$ and $b_1/b,\ldots,b_b/b$, respectively. Apply lemma 5 to complete the proof.

Lemma 7. *Let* $a_1,\ldots,a_n$ *and* $c_1,\ldots,c_n$ *be two sequences of* n *real numbers with*

$$(2.11) \qquad \sum_1^n c_i = 0 \quad \text{and} \quad \sum_1^n |c_i| = 1$$

for $n \geq 2$. *Let* Φ *be the set of permutations of the indices* 1 *through* n *with typical member* $\varphi = (\varphi_1,\ldots,\varphi_n)$. *Then*

$$(2.12) \qquad \max(a_1,\ldots,a_n) \le \frac{1}{n}\sum_1^n a_i + \frac{1}{n(n-2)!}\sum_{\varphi\in\Phi}^{n} \left|\sum_{i=1}^n c_{\varphi_i} a_i\right| .$$

Proof. If we replace each summand $\left|\sum_{i=1}^n c_{\varphi_i} a_i\right|$ by $\sum_1^n c_{\varphi_i} a_i$ when $c_{\varphi_1} \ge 0$ and by $-\sum_1^n c_{\varphi_i} a_i$ otherwise, then, the right hand side (R.H.S.) of (2.12) simplifies to a_1. This replacement process can do nothing more than reduce the right hand side. Thus, $a_1 \le$ R.H.S. The proof is completed by applying the same argument for all indices 1 through n.

The next lemma has its roots in game theory.

Lemma 8. *Let* $M(x,y)$ *be a continuous function over the domain* $X \times Y$, $x\in X$ *and* $y\in Y$, *where* X *and* Y *are compact, convex regions in finite dimensional Euclidean spaces. Suppose that* M *is a convex function in* y *for each* x *and a concave function in* x *for each* y. *Then*

$$(2.13) \qquad \min_{y\in Y}\max_{x\in X} M(x,y) = \max_{x\in X}\min_{y\in Y} M(x,y) .$$

(This lemma immediately implies the fundamental theorem for rectangular games.)

Proof. The lemma may be easily verified using a theorem (theorem 2) given in a joint paper by Bohnenblust, Karlin, and Shapley [4].

3. Generalized Wald lower bound for ASN.

This section uses lemmas 1, 3, 4, and 5.

In the next several sections we will find it convenient to interpret $0 \ln 0/c$ as 0 for $c \geq 0$ and $c \ln c/0$ as ∞ for $c > 0$.

Theorem 3.1. [1/] (Generalized Wald lower bound for ASN) *Let $X_1, X_2, \ldots$ be a sequence of independent random variables identically distributed as X. Let $H_1, \ldots, H_K$ be K hypotheses where H_i is the hypothesis that X has density function f_i with respect to some measure μ, for $i = 1, \ldots, K$, $K \geq 2$. Assume that f_i and f_j are not identical a.e. (μ) for $i \neq j$. Let N be the number of observations in a sequential test (randomized or not randomized) which chooses one of the K densities subject to a $K \times K$ error matrix $A = (\alpha_{ij})$ where $\alpha_{ij} = P_i[$ accepting $H_j]$. For given index i, assume that $\alpha_{i\nu} = 0$ whenever any $\alpha_{j\nu} = 0$. Then a lower bound for $E_i(N)$ is given by*

1/ Although theorem 3.1 is due to the author, a similar unpublished theorem due to W. Hoeffding was found to exist some time after the author's discovery. His theorem will appear in print "by permission" in Sequential Procedures for Ranking and Identification Problems, University of Chicago Statistics Monograph Series, by R.E. Beckhofer, J. Kiefer, and M. Sobel.

$$(3.1) \qquad \max_{\substack{1 \leq j \leq K \\ j \neq i}} \frac{\sum_{\nu=1}^{K} \alpha_{i\nu} \ln \frac{\alpha_{i\nu}}{\alpha_{j\nu}}}{\int f_i \ln(f_i/f_j) d\mu} \quad \underline{\text{for}} \quad i=1,\ldots,K .$$

Proof. Let i and j be fixed distinct indices between 1 and K. $\sum_{m=1}^{n} \ln \frac{f_i(X_m)}{f_j(X_m)}$ is the sum of n independent random variables identically distributed as $\ln \frac{f_i(X)}{f_j(X)}$. We may assume that $E_i(N) < \infty$. Otherwise (3.1) is trivially a lower bound. Suppose for now that

$$(3.2) \qquad E_i(\ln \frac{f_i(X)}{f_j(X)}) = \int f_i \ln(f_i/f_j) d\mu < \infty .$$

Then, by lemma 4, $E_i(\ln \frac{f_i(X)}{f_j(X)})$ is finite. Lemma 1 yields

$$(3.3) \qquad E_i(\sum_{m=1}^{N} \ln \frac{f_i(X_m)}{f_j(X_m)}) = E_i(N) E_i(\ln \frac{f_i(X)}{f_j(X)}) .$$

Let $D = \{\nu | \alpha_{i\nu} \neq 0, \ \nu = 1,\ldots,K\}$. In accordance with the notation of lemma 3, we can write

$$(3.4) \qquad E_i(\sum_{m=1}^{N} \ln \frac{f_i(X_m)}{f_j(X_m)}) = \sum_{\nu \in D} \alpha_{i\nu} E_i^{\nu} (\sum_{m=1}^{N} \ln \frac{f_i(X_m)}{f_j(X_m)}) ,$$

by breaking up the sample space (possibly randomized) into the regions on which the various hypotheses are accepted. Applying the conditional Jensen's inequality to the continuous convex function $-\ln(x)$, yields

$$(3.5) \qquad E_i^{\nu}(\sum_{m=1}^{N} \ln \frac{f_i(X_m)}{f_j(X_m)}) \geq -\ln E_i^{\nu}(\prod_{m=1}^{N} \frac{f_j(X_m)}{f_i(X_m)}) .$$

But, by lemma 3,

$$(3.6) \qquad E_i^\nu\left(\prod_{m=1}^{N} \frac{f_j(X_m)}{f_i(X_m)}\right) \leq \frac{\alpha_{j\nu}}{\alpha_{i\nu}} ,$$

Combining (3.3), (3.4), (3.5), and (3.6), we get

$$(3.7) \qquad E_i(N)E_i\left(\ln \frac{f_i(X)}{f_j(X)}\right) \geq \sum_{\nu \in D} \alpha_{i\nu} \ln \frac{\alpha_{i\nu}}{\alpha_{j\nu}} .$$

The convention which interprets $\alpha_{i\nu} \ln \frac{\alpha_{i\nu}}{\alpha_{j\nu}}$ as zero when $\alpha_{i\nu} = 0$ allows us to rewrite (3.7) as

$$(3.8) \qquad E_i(N)E_i\left(\ln \frac{f_i(X)}{f_j(X)}\right) \geq \sum_{\nu=1}^{K} \alpha_{i\nu} \ln \frac{\alpha_{i\nu}}{\alpha_{j\nu}} .$$

Thus

$$(3.9) \qquad E_i(N) \geq \frac{\sum_{\nu=1}^{K} \alpha_{i\nu} \ln \frac{\alpha_{i\nu}}{\alpha_{j\nu}}}{E_i\left(\ln \frac{f_i(X)}{f_j(X)}\right)} = \frac{\sum_{\nu=1}^{K} \alpha_{i\nu} \ln \frac{\alpha_{i\nu}}{\alpha_{j\nu}}}{\int f_i \ln(f_i/f_j)d\mu} .$$

Now, even if the inequality of (3.2) does not hold, (3.9) still holds trivially. Since index j is arbitrary except for $j \neq i$, the theorem follows.

Theorem 3.1 allows us to find a lower bound for ASN when one of the hypotheses is true. For $K = 2$, the lower bound is identical to the lower bound given by Wald [14]. The next section finds one lower bound for ASN when none of the hypotheses are true.

4. A first lower bound for ASN when none of the K hypotheses is true.

This section uses lemmas 5 and 8, and theorem 3.1.

Theorem 4.1. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables identically distributed as X. Consider any test of hypotheses where we are to choose among K densities f_i (with respect to some measure μ), $i = 1, \ldots, K$; $K \geq 2$. Let $A = (\alpha_{ij})$ be the $K \times K$ error matrix with $\alpha_{ij} = P_i$ (accepting f_j). Assume $\alpha_{ij} > 0$ for $i, j = 1, \ldots, K$. Let f_0 be a $K+1^{st}$ density (with respect to μ). Let N be the (random) number of observations in the test. Then a lower bound for $E_0(N)$ is given by*

$$\inf_{\{b_\nu\}} \max_{1 \leq j \leq K} \frac{\sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_j)d\mu}, \tag{4.1}$$

where

$$b_\nu > 0 \quad \text{for} \quad \nu = 1, \ldots, K \quad \text{and} \quad \sum_1^K b_\nu = 1. \tag{4.2}$$

Proof. Let T be a sequential test satisfying error matrix A. Let $b_\nu = P_0[\text{accepting } f_\nu]$, $\nu = 1, \ldots, K$. (For the present, we allow b_ν to be zero.) Interpreting f_0 as a $K+1^{st}$ "hypothesis" we have a $(K+1) \times (K+1)$ error matrix

$$A^* = \left(\begin{array}{c|ccc} 0 & b_1 & \cdots & b_K \\ \hline 0 & & & \\ \vdots & & A & \\ 0 & & & \end{array}\right). \tag{4.3}$$

Using theorem 3.1, we may conclude that among all tests $T = T(\{b_\nu\})$ which satisfy A^* (and hence A),

$$(4.4) \qquad E_0(N) \geq \max_{1 \leq j \leq K} \frac{\sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_j) d\mu} .$$

For a different set $\{b_\nu'\}$, we get a different bound

$$(4.5) \qquad E_0(N) \geq \max_{1 \leq j \leq K} \frac{\sum_{\nu=1}^{K} b_\nu' \ln(b_\nu'/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_j) d\mu} .$$

For any test of the type $T(\{b_\nu\})$ or $T(\{b_\nu'\})$, $E_0(N)$ is bounded by the minimum of the two bounds in (4.4) and (4.5). Since we must consider every test which satisfies A regardless of the values of the set $\{b_\nu\}$, it follows that

$$(4.6) \qquad E_0(N) \geq \inf_{\{b_\nu\}} \max_{1 \leq j \leq K} \frac{\sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_j) d\mu} ,$$

where $b_\nu \geq 0$ and $\sum_1^K b_\nu = 1$. The same bound results if we insist that $b_\nu > 0$ and $\sum_1^K b_\nu = 1$. This completes the proof.

In the proof above, it is reasonable to ask whether it is necessary to take the infimum over so large a class of sets. Might there be certain sets $\{b_\nu\}$ for which no test T exists such that $b_\nu = P_0$ [accepting f_ν] for $\nu = 1,\ldots,K$? A reduction in the size of the class of sets that we take the infimum over might make the lower bound for ASN larger. The following theorem shown that we cannot reduce the class size to advantage under most situations.

Theorem 4.2. Let $A = (\alpha_{ij})$ be a $K \times K$ error matrix where each column of A is composed of non-zero elements or composed of only zero elements. Assume there exists a sequence of tests $T_1, T_2, \ldots$ with error matrices $B_1, B_2, \ldots$ respectively for which $\lim_{n \to \infty} B_n = I_K$ (the $K \times K$ identity matrix). Equivalently, we assume the existence of a consistent sequence of tests. Then there exists a randomized test with error matrix A.

Proof. We will assume as obvious that, since $\lim_{n\to\infty} B_n = I_K$, there exists as N such that, for $n \geq N$, B_n^{-1} exists and $\lim_{N \leq n\to\infty} B_n^{-1} = I_K$. Without loss of generality, assume that B_n^{-1} exists for all $n \geq 1$. For any stochastic matrix P, we can produce a test T_n' with error matrix $B_n P$. One simply modifies test T_n by accepting hypothesis H_ν with probability $p_{i\nu}$ when T_n says to accept H_i. It suffices to find a positive index n and stochastic matrix P for which $A = B_n P$. But $P_n \equiv B_n^{-1} A \to A$ as $n \to \infty$. From the assumption that each column of A has no zero elements or only zero elements, we conclude that the general element of P_n, namely p_{nij}, is zero for all n or approaches a positive limit. It follows that for sufficiently large n, we can define $P \equiv P_n$.

The lower bound given in theorem 4.1 can be written in an alternative form which, for $K = 2$, is identical to a lower bound given by Hoeffding [7].

Theorem 4.3. (Equality of two lower bounds for ASN) Let $0 < \int f_0 \ln(f_0/f_j)d\mu < \infty$ for $j = 1,\ldots,K$. The following two expressions are equal:

(4.6)
$$\inf_{\{b_\nu\}} \max_{1 \le j \le K} \frac{\sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_j)d\mu}$$

<u>and</u>

(4.7)
$$\sup_{\{c_j\}} \frac{-\ln\{\sum_{\nu=1}^{K} \prod_{j=1}^{K} \alpha_{j\nu}^{c_j}\}}{\sum_{j=1}^{K} c_j \int f_0 \ln(f_0/f_j)d\mu},$$

<u>where</u>

(4.8) $\sum_{\nu=1}^{K} b_\nu = \sum_{j=1}^{K} c_j = 1$; $b_\nu > 0$, <u>for</u> $\nu = 1,\ldots,K$; <u>and</u> $c_j \ge 0$ <u>for</u> $j=1,\ldots,K$,

<u>and</u>

(4.9) $\sum_{\nu=1}^{K} \alpha_{j\nu} = 1$ <u>for</u> $j=1,\ldots,K$; $\alpha_{ij} > 0$ <u>for</u> $i, j=1,\ldots,K$.

<u>Proof</u>. For abbreviation, let $s_j \equiv \int f_0 \ln(f_0/f_j)d\mu$ and, for $b = (b_1,\ldots,b_k)$,

(4.10)
$$h_j(b) \equiv \sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu}) .$$

The argument, which uses lemmas 5 and 8 proceeds through a series of equalities:

$$\inf_{\{b_\nu\}} \max_{1 \le j \le K} h_j(b)/s_j$$

(4.11a)
$$= \inf_{\{b_\nu\}} \sup_{\{c_j\}} \left\{\sum_{j=1}^{K} c_j h_j(b) / \sum_{j=1}^{K} c_j s_j\right\}$$

(4.11b)
$$= \inf_{\{b_\nu\}} \sup_{\{c_j^*\}} \sum_{j=1}^{K} c_j^* h_j^*(b)$$

$$(4.11c) \qquad = \sup_{\{c_j^*\}} \inf_{\{b_\nu\}} \sum_{j=1}^{K} c_j^* h_j^*(b)$$

$$(4.11d) \qquad = \sup_{\{c_j\}} \inf_{\{b_\nu\}} \left\{ \sum_{j=1}^{K} c_j h_j(b) / \sum_{j=1}^{K} c_j s_j \right\}$$

$$(4.11e) \qquad = \sup_{\{c_j\}} \left\{ -\ln[\sum_{\nu=1}^{K} \prod_{j=1}^{K} \alpha_{j\nu}^{c_j}] / \sum_{j=1}^{K} c_j s_j \right\} ,$$

where

$$c_j^* \equiv c_j s_j / \sum_{i=1}^{K} c_i s_i , \quad \text{and} \quad h_j^*(b) \equiv h_j(b)/s_j , \quad \text{for} \quad j=1,\dots,K .$$

The set $\{c_j^*\}$ satisfies the same requirements as $\{c_j\}$ does in (4.8).

Equality (4.11a) follows immediately from the obvious equality:

$$\max_{1 \le j \le K} h_j(b)/s_j = \sup_{\{c_j\}} \left\{ \sum_{j=1}^{K} c_j h_j(b) / \sum_{j=1}^{K} c_j s_j \right\} .$$

(Note: $0 < s_j < \infty$ by assumption, while $h_j(b) \ge 0$ because of lemma 5.)

Equalities (4.11b) and (4.11d) are a consequence of the equivalence between taking supremums with respect to $\{c_j\}$ and taking them with respect to $\{c_j^*\}$. Equality (4.11c) is verified by using lemma 8. The application requires us to extend the definition of $h_j^*(b)$ to the boundary of its domain. This may be done by insisting on continuity.

Finally, equality (4.11e) follows upon verification of

$$(4.12) \qquad \inf_{\{b_\nu\}} G(b;c) = -\ln \sum_{\nu=1}^{K} \prod_{j=1}^{K} \alpha_{j\nu}^{c_j}$$

where, for $c = (c_1,\dots,c_K)$, we define

$$(4.13) \qquad G(b;c) \equiv \sum_{j=1}^{K} c_j h_j(b) .$$

We shall need the first and second partial derivatives of h_j. Because of (4.8), we shall interpret $h_j(b)$ implicitly as a function of its 1^{st} K-1 coordinates. Then

(4.14) $$h_{j\nu}(b) = \ln(b_\nu \alpha_{jK}/b_K \alpha_{j\nu}) \ , \quad \text{for} \quad \nu=1,\ldots,K-1 \ ,$$

and

$$h_{j\nu\nu'}(b) = \delta_{\nu\nu'}/b_K + 1/b_K \ , \quad \text{for} \quad \nu=1,\ldots,K-1$$

where $\delta_{\nu\nu'}$ is the Kronecker delta. From

$$\sum_{\nu,\nu'=1}^{K-1} h_{j\nu\nu'} b_\nu b_{\nu'} = (1-b_K)/b_K > 0,$$

we conclude that $h_j(b)$ is convex.

It follows that $G(b;c)$ is convex in b and any solution of $G_\nu \equiv \frac{\partial G}{\partial b_\nu} = 0$, $\nu=1,\ldots,K-1$, must correspond to the minimum of G. Setting $G_\nu = 0$, we find

(4.15) $$\sum_{j=1}^{K} c_j \ln(b_\nu \alpha_{jK}/b_K \alpha_{j\nu}) = 0 \ , \quad \text{for} \quad \nu=1,\ldots,K-1 \ .$$

(4.15) also holds for $\nu = K$. Thus

(4.16) $$b_\nu = b_K \prod_{j=1}^{K} \alpha_{j\nu}^{c_j} / \prod_{j=1}^{K} \alpha_{jK}^{c_j} \quad \text{for} \quad \nu=1,\ldots,K \ ,$$

and so (summing over ν)

(4.17) $$1 = b_K \sum_{\nu=1}^{K} \prod_{j=1}^{K} \alpha_{j\nu}^{c_j} / \prod_{j=1}^{K} \alpha_{jK}^{c_j}$$

(4.16) and (4.17) verify the existence of a solution b (satisfying (4.8)).

Now, multiply (4.15) by b_ν, sum over ν, and simplify to show that

(4.18) $$G(b;c) = \ln(b_K / \prod_{j=1}^{K} \alpha_{jK}^{c_j}) \ .$$

Equality (4.12) (and hence (4.11e)) follows from (4.17) and (4.18).

As a rule, expression (4.6) is easier to compute than expression (4.7) because $h_j(b)$ is convex in b for each j.

5. A second lower bound for ASN when none of the K hypotheses is true.

This section uses lemma 1, 2, 4, and 7. Theorem 5.1 below gives Hoeffding's bound (1.5) for ASN when $K = 2$ and is a generalization for $K > 2$. Three of Hoeffding's regularity conditions are replaced by (5.1) below, thus avoiding a condition which depends on the sequential test under study and a fourth assumption is unnecessary.

Theorem 5.1. *Let* $X_1, X_2, \ldots$ *be a sequence of independent random variables identically distributed as* X. *Consider any test of hypotheses where we are to choose among* K *densities* f_i *(with respect to some measure* μ*),* $i=1,\ldots,K$; $K \geq 2$. *Let* $A = (\alpha_{ij})$ *be the* $K \times K$ *error matrix with* $\alpha_{ij} = P_i[\text{accepting } f_j]$. *We assume* $\alpha_{ij} > 0$ *for* $i, j=1,\ldots,K$. *Let* f_0 *be a* $(K+1)^{st}$ *density (with respect to* μ*). Assume that*

$$(5.1) \qquad \int f_0 \ln^2(f_0/f_i)\,d\mu < \infty \ , \quad \text{for} \quad i=1,\ldots,K \ .$$

Let

$$(5.2) \qquad D = \{i_1,\ldots,i_\nu\}$$

be a subset of the first K positive integers with ν distinct members, $\nu \geq 2$. Let N be the (random) number of observations in the test. Then

$$E_0(N) \geq [(T^2(D) - R(D)\ln S(D))^{\frac{1}{2}} - T(D)]^2/R^2(D) , \tag{5.3}$$

where

$$R(D) = \max_{i\epsilon D} \int f_0 \ln(f_0/f_i)d\mu , \tag{5.4}$$

$$S(D) = \sum_{j=1}^{K} \min_{i\epsilon D} \alpha_{ij} , \tag{5.5}$$

and where $T(D)$ is defined in the following manner. Let $C = \{c_i | i\epsilon D\}$ be any set of ν real numbers for which

$$\sum_{i\epsilon D} c_i = 0 \quad \text{and} \quad \sum_{i\epsilon D} |c_i| = 1. \tag{5.6}$$

Let $\Phi(D)$ be the permutations of D with typical member $\varphi = (\varphi_i ; i\epsilon D)$, a ν-dimensional vector. Then

$$T(D) = \inf_{C} \tau(C)/2\nu(\nu-2)! , \tag{5.7}$$

where

$$\tau(C) = \sum_{\varphi\epsilon\Phi(D)} \tau_\varphi(C) , \tag{5.8}$$

and where

$$(5.9)\qquad \tau_\varphi^2(C) = \int f_0\Big(\sum_{i\epsilon D} c_{\varphi_i}[\ln(f_0/f_i) - E_0(\ln(f_0/f_i))]\Big)^2 d\mu \ .$$

Equivalently, $\tau_\varphi^2(C)$ is the variance under f_0 of $\sum_{i\epsilon D} c_{\varphi_i} \ln(f_0/f_i)$.

Remark. For each subset D, we have a different lower bound given by (5.3). Obviously, one is interested in the largest lower bound one can obtain by considering various sets D.

Proof. Adopting the same notation as that used in proving lemma 3, we represent the test notationally by (ψ, ϕ). Then

$$\alpha_{ij} = E_i(\phi_{Nj}) = \psi_0\phi_{0j} + \sum_{n=1}^{\infty} \int \psi_n \phi_{nj} f_{in} d\mu^n \ ,$$

where

$$(5.10)\qquad f_{in} = \prod_{m=1}^{n} f_i(X_m) \ ,$$

and μ^n is the n-fold product measure generated by μ. Using $\sum_{j=1}^{K} \phi_{nj} = 1$, for $n = 0,1,\ldots,$ we get

$$(5.11)\qquad \begin{aligned} S(D) &= \sum_{j=1}^{K} \min_{i\epsilon D} \alpha_{ij} = \psi_0 \sum_{j=1}^{K} \phi_{0j} + \sum_{j=1}^{K} \min_{i\epsilon D} \sum_{n=1}^{\infty} \int \psi_n \phi_{nj} f_{in} d\mu^n \\ &\geq \psi_0 + \sum_{j=1}^{K} \sum_{n=1}^{\infty} \int \psi_n \phi_{nj} \cdot \min_{i\epsilon D} f_{in} \cdot d\mu^n = \psi_0 + \sum_{n=1}^{\infty} \int \psi_n \cdot \min_{i\epsilon D} f_{in} \cdot d\mu^n . \end{aligned}$$

But

$$(5.12)\qquad \int \psi_n \cdot \min_{i\epsilon D} f_{in} \cdot d\mu^n \geq \int_{A_n} \psi_n \cdot \min_{i\epsilon D} \frac{f_{in}}{f_{0n}} \cdot f_{0n} d\mu^n \ ,$$

where f_{0n} is defined as in (5.10), and $A_n = [f_{0n} > 0]$. Defining $\frac{f_{i0}}{f_{00}} \equiv 1$ and combining (5.11) and (5.12), we find that

$$S(D) \geq \psi_0 \cdot \min_{i\epsilon D} \frac{f_{i0}}{f_{00}} + \sum_{n=1}^{\infty} \int_{A_n} \psi_n \cdot \min_{i\epsilon D} \frac{f_{in}}{f_{0n}} \cdot f_{0n} d\mu^n$$

(5.13)

$$= E_0(\min_{i\epsilon D} \frac{f_{iN}}{f_{0N}}) \; .$$

Now define

(5.14) $$Z_{ni} \equiv \sum_{m=1}^{n} \{\ln \frac{f_0(X_m)}{f_i(X_m)} - E_0(\ln \frac{f_0}{f_i})\} \; .$$

Then

(5.15) $$Z_{ni} = \ln \frac{f_{0n}}{f_{in}} - n\zeta_i \; ,$$

where $\zeta_i = \int f_0 \ln(f_0/f_i) d\mu$. ζ_i is finite because of regularity condition (5.1) which in turn implies the almost sure finiteness of Z_{ni} with respect to density f_0. It follows (from (5.4),(5.13), (5.15), and Jensen's inequality) that

$$S(D) \geq E_0(\min_{i\epsilon D} \frac{f_{iN}}{f_{0N}}) = E_0(\exp\{-\max_{i\epsilon D}(Z_{Ni} + N\zeta_i)\})$$

(5.16)

$$\geq E_0(\exp\{-\max_{i\epsilon D} Z_{Ni} - N \max_{i\epsilon D} \zeta_i\}) = E_0(\exp\{-\max_{i\epsilon D} Z_{Ni} - NR(D)\})$$

$$\geq \exp \{-E_0(\max_{i\epsilon D} Z_{Ni}) - E_0(N)R(D)\}.$$

Thus

(5.17) $$\ln S(D) \geq -E_0(\max_{i\epsilon D} Z_{Ni}) - E_0(N)R(D) \; .$$

Lemma 7 gives us

$$\max_{i\in D} Z_{Ni} \le \frac{1}{\nu} \sum_{i\in D} Z_{Ni} + \frac{1}{\nu(\nu-2)!} \sum_{\varphi\in\Phi(D)} \left| \sum_{i\in D} c_{\varphi_i} Z_{N_i} \right| ,$$

and hence,

$$(5.18) \qquad E_0(\max_{i\in D} Z_{Ni}) \le \frac{1}{\nu} \sum_{i\in D} E_0(Z_{Ni}) + \frac{1}{\nu(\nu-2)!} \sum_{\varphi\in\Phi(D)} E_0 \left| \sum_{i\in D} c_{\varphi_i} Z_{Ni} \right| .$$

Now, if $E_0(N) = \infty$, the lower bound is trivial. So, let us assume that

$$(5.19) \qquad E_0(N) < \infty .$$

In (5.14), Z_{ni} is defined to be the sum of n independent and identically distributed random variables, each with mean zero. Thus, under assumption (5.19), lemma 1 applies to show $E_0(Z_{Ni}) = 0$ for $i\in D$, and (5.18) simplifies to

$$(5.20) \qquad E_0(\max_{i\in D} Z_{Ni}) \le \frac{1}{\nu(\nu-2)!} \sum_{\varphi\in\Phi(D)} E_0 \left| \sum_{i\in D} c_{\varphi_i} Z_{Ni} \right| .$$

Assumption (5.1) implies $E_0(Z_{1i}^2) < \infty$ which, coupled with Schwartz' inequality, leads to

$$(5.21) \qquad E_0(\sum_{i\in D} c_{\varphi_i} Z_{1i})^2 \le \sum_{i\in D} c_{\varphi_i}^2 \sum_{i\in D} E_0(Z_{1i}^2) < \infty .$$

Also,

$$(5.22) \qquad E_0(\sum_{i\in D} c_{\varphi_i} Z_{1i}) = 0 .$$

$\sum_{i\in D} c_{\varphi_i} Z_{ni}$ is the sum of n independent random variables identically distributed as $\sum_{i\in D} c_{\varphi_i} Z_{1i}$ and lemma 2 applies. (Use (5.18), (5.21), and (5.22).) Hence,

$$E_0^2\left|\sum_{i\in D} c_{\varphi_i} Z_{Ni}\right| \le E_0\left(\sum_{i\in D} c_{\varphi_i} Z_{Ni}\right)^2 = E_0(N)E_0\left(\sum_{i\in D} c_{\varphi_i} Z_{1i}\right)^2$$

(5.23)

$$= E_0(N)\ \tau_\varphi^2(C).$$

Combining (5.20) and (5.23),

$$E_0(\max_{i\in D} Z_{Ni}) \le \frac{E_0^{\frac{1}{2}}(N)}{\nu(\nu-2)!} \sum_{\varphi\in\Phi(D)} \tau_\varphi(C)$$

$$= \frac{E_0^{\frac{1}{2}}(N)}{\nu(\nu-2)!}\ \tau(C)\ .$$

Since the set C was chosen arbitrarily, we may take the infimum of $\tau(C)$ over all sets C. Then

$$E_0(\max_{i\in D} Z_{Ni}) \le \frac{E_0^{\frac{1}{2}}(N)}{\nu(\nu-2)!} \inf_C \tau(C) = 2E_0^{\frac{1}{2}}(N)T(D)\ .$$

Returning to (5.17), we can form the quadratic inequality in $E_0^{\frac{1}{2}}(N)$ as

(5.24) $$\ln S(D) \ge -2E_0^{\frac{1}{2}}(N)T(D) - E_0(N)R(D)\ .$$

Its solution provides the lower bound (5.3) for $E_0(N)$. ($R(D) > 0$, because of lemma 4.) This completes the proof.

Even though theorem 5.1 appears rather complicated, it may be computationally easier to apply than theorem 4.1. Consider the following example.

Example. Let $f_0, f_1, \ldots, f_K$ be normal densities with means $\theta_0, \theta_1, \ldots, \theta_K$, respectively, and with common variance σ^2. Then (5.4) becomes

$$(5.25)\qquad R(D) = \max_{i\in D} \int f_0 \ln(f_0/f_i)d\mu = \frac{1}{2\sigma^2}\max_{i\in D}(\theta_0-\theta_i)^2 ,$$

and (5.7) becomes

$$(5.26)\qquad T(D) = \frac{1}{2\nu(\nu-2)!} \inf_C \sum_{\varphi\in\Phi(D)} \left|\sum_{i\in D} c_{\varphi_i} \frac{\theta_i}{\sigma}\right| .$$

It is interesting to compare (5.26) with (2.12) of lemma 7. The sum, $\frac{1}{n(n-2)!}\sum_{\varphi\in\Phi}\left|\sum_{i=1}^{n} c_{\varphi_i} a_i\right|$, is quite similar to the sum in (5.26).

The infimum of (5.26) is easily handled when $\nu = 2$ or 3 (and thus for $K = 2$ or 3). The infimum for $\nu = 2$ is achieved with $C = \{\frac{1}{2}, -\frac{1}{2}\}$, and for $\nu = 3$ is achieved with $C = \{\frac{1}{4}, \frac{1}{4}, -\frac{1}{2}\}$. Then, for $\nu = 2$, $T(D) = \frac{1}{4\sigma}|\theta_{i_1}-\theta_{i_2}|$, where $i_1, i_2 \in D$, $i_1 \neq i_2$; and, for $\nu = 3$,

$$(5.27)\qquad T(D) = \frac{1}{4\sigma}(\theta_{max}-\theta_{min}) + \frac{1}{12\sigma}|\theta_{max} + \theta_{min} - 2\theta_{mid}| ,$$

where θ_{min}, θ_{mid}, and θ_{max} is the ordering of θ_{i_1}, θ_{i_2}, and θ_{i_3}, and where i_1, i_2, and i_3 are the 3 distinct integers in D.

It is unknown to the author whether "universal minimizing" sets C exist for $\nu \geq 4$, and if so, what they are.

Theorem 5.1 assumes very little concerning the nature of the set of densities $f_0, f_1, \ldots, f_K$. If either one of two frequently satisfied assumptions is valid, we can improve upon the theorem. We will need a few definitions. A set of real valued functions $\{g_i(X);\ i\in I\}$ will be said to be <u>pairwise minimizable</u> if, for every finite subset $D \subset I$ with two or more indices, there is a subset $D' \subset D$ with two members such that $\min_{i\in D} g_i(X) = \min_{i\in D'} g_i(X)$ for all x. The two functions

$g_i(X)$, $i\epsilon D'$ will be referred to as the minimizing functions. A doubly indexed set of real valued functions $\{g_{ij}(x_j)\colon i\epsilon I,\ j\epsilon J\}$ will be said to be uniformly pairwise minimizable in i if $\{g_{ij}(x_j)\colon i\epsilon I\}$ is pairwise minimizable for all $j\epsilon J$ and if the sets D' do not depend on j.

Now, consider the following two conditions:

C_1: $\{f_{in}\colon i=1,\ldots,K;\ n=1,2,\ldots\}$ is uniformly pairwise minimizable in i.

C_2: $\{\ln f_{in}-E_0(\ln f_{in})\colon i=1,\ldots,K;\ n=1,2,\ldots\}$ is uniformly pairwise minimizable in i.

It can be shown that C_1 holds whenever $f_1,\ldots,f_K$ are members of the same exponential family of the form $c(\theta)\,h(x)\,\exp(Q(\theta)t(x))$, and C_2 holds if, in addition, we have $E_0|\ln f_i| < \infty$ for $i=1,\ldots,K$. Verification of C_2 is direct while C_1 follows from the fact that $\ln[c(\theta)h(x)\exp(Q(\theta)t(x))]$ is a concave function in $Q(\theta)$. In both cases, the minimizing functions are associated with the densities which have the smallest and largest value of $Q(\theta)$.

If C_1 holds, we can replace the expression $\min\limits_{i\epsilon D} \frac{f_{iN}}{f_{ON}}$ in (5.16) by $\min\limits_{i\epsilon D'} \frac{f_{iN}}{f_{ON}}$. If C_2 holds, we can replace the expression $-\max\limits_{i\epsilon D} Z_{Ni}$ in (5.16) by $-\max\limits_{i\epsilon D'} Z_{Ni}$. Theorem 5.1 is modified to the extent that under C_1 we must redefine

$$R(D) = \max_{i\epsilon D'} \int f_0 \ln(f_0/f_i)\,d\mu \quad , \tag{5.28}$$

and

$$T^2(D) = \frac{1}{16} \int [\ln(f_i/f_j)-E_0(\ln(f_i/f_j))]^2 d\mu \tag{5.29}$$

for $i,\ j\epsilon D'$, $i \neq j$.

If just C_2 holds, we only redefine $T(D)$ using (5.29).

In the example above with normal densities, $R(D)$ is not really changed (see (5.25)), but $T(D)$ is improved for $\nu = 3$ whenever $|\theta_{max} + \theta_{min} - 2\theta_{mid}| \neq 0$ (see (5.27)).

Remark. Note that at no point in the proof of Theorem 5.1 did we require f_0 to differ from the set $f_1,\ldots,f_K$. Thus theorem 5.1 applies when one of the hypotheses is true, also.

6. Lower bounds for ASN subject to partial control of the error matrix.

This section uses lemma 6 and Theorems 3.1, 4.1, and 5.1. We will use the notation

$$\hat{Q} \equiv 1-Q \ , \tag{6.1}$$

where Q is used in a generic sense.

To this point we have treated the error matrix as completely fixed. This is not satisfactory in most applications. Nevertheless, there seems to be some merit in introducing a set of techniques by applying them to a specific problem. More importantly, the previous results, either directly or by analogy, provide us with lower bounds for ASN under a rather wide variety of situations. In this section, we will consider problems in which the control of the error matrix is relaxed or modified.

Consider the following concrete example. The author [11] has investigated a three hypothesis sequential test for the unknown mean of a normal distribution in which one can readily control the 3×3 error matrix in one of these two ways:

(i) Fix just the main diagonal.

(ii) Fix the main diagonal and second row.

If f_0 is different from the other three densities f_1, f_2, and f_3, we can start with expression (4.1) of Theorem 4.1:

$$(6.2) \qquad \inf_{\{b_\nu\}} \max_{1 \leq j \leq 3} \frac{\sum_{\nu=1}^{K} b_\nu \ln(b_\nu/\alpha_{j\nu})}{\int f_0 \ln(f_0/f_1)}$$

where $b_1 + b_2 + b_3 = 1$; $b_\nu > 0$, $\nu = 1,2,3$. In a manner completely analogous to the method used in proving Theorem 4.1, we find lower bounds for case (i) and (ii) by taking infimums to get rid of "over-controls". In case (i) we take an infimum of (6.2) with respect to the elements α_{ij} off the diagonal and in case (ii) by taking an infimum of (6.2) with respect to the α_{ij} off the main diagonal and out of the second row. In both cases the infimum with respect to the α_{ij} can be interchanged with the expression "$\inf_{\{b_\nu\}} \max_{1\leq j\leq 3}$". The result is that we must compute quantities such as

$$\inf_{\{\alpha_{12},\alpha_{13}:\ \sum_{j=1}^{3}\alpha_{1j} = 1;\ \alpha_{11},\alpha_{12},\alpha_{13}>0\}} \sum_{\nu=1}^{3} b_\nu \ln b_\nu/\alpha_{1\nu}.$$

Lemma 6 tells us that the infimum is equal to

$$b_1 \ln(b_1/\alpha_{11}) + \hat{b}_1 \ln \hat{b}_1/\hat{\alpha}_{11} .$$

Thus a lower bound for $E_0(N)$ is given for situation (i) by

$$(6.3) \qquad \inf_{\{b_j\}} \max_{1 \leq j \leq 3} \frac{b_j \ln(b_j/\alpha_{jj}) + \hat{b}_j \ln(\hat{b}_j/\hat{\alpha}_{jj})}{\int f_0 \ln(f_0/f_j) d\mu}$$

and for situation (ii) by

$$(6.4)\qquad \inf_{\{b_\nu\}} \max\left\{\frac{b_1\ln(b_1/\alpha_{11})+\hat{b}_1\ln(\hat{b}_1/\hat{\alpha}_{11})}{\int f_0\ln(f_0/f_1)d\mu}\,,\ \frac{\sum_{\nu=1}^{3} b_\nu\ln(b_\nu/\alpha_{2\nu})}{\int f_0\ln(f_0/f_2)d\mu}\,,\right.$$

$$\left.\frac{b_3\ln(b_3/\alpha_{33})+\hat{b}_3\ln(\hat{b}_3/\hat{\alpha}_{33})}{\int f_0\ \ln(f_0/f_3)d\mu}\right\}$$

where

$$(6.5)\qquad b_1+b_2+b_3=1 \ \text{ and } \ b_j>0 \ \text{ for } \ j=1,2,3\ .$$

Remark 1. The method used in deriving (6.3) and (6.4) illustrates a general method whereby lower bounds for ASN can be "made to order" based on the constraint of any portion of the error matrix. It is clearly possible to consider problems with more complicated constraints such as the constraint of the sum or the maximum of the α_{ij}'s which are off the main diagonal.

Remark 2. Let $B(D)$ be the lower bound for ASN for the set D as given by theorem 5.1. Let $\mathcal{D}$ be the class of all such sets D. Then, of course,

$$E_0(N) \geq \max_{D\in\mathcal{D}} B(D)\ .$$

This bound, which was derived for tests constraining the entire error matrix, can be used to find lower bounds for situations (i) and (ii) above. One need only take the appropriate infimums.

When one of the hypotheses is true, cases (i) and (ii) are slightly more difficult but one can show (using (3.1) of Theorem 3.1) that in case (i) a lower bound for $E_i(N)$ is given by

$$(6.6) \qquad \inf_{\{\alpha_{ij}:\ j\neq i\}} \ \max_{\substack{1\leq j\leq 3 \\ j\neq i}} \ \frac{\alpha_{ij}\ln(\alpha_{ij}/\alpha_{jj})+\hat{\alpha}_{ij}\ln(\hat{\alpha}_{ij}/\hat{\alpha}_{jj})}{\int f_i \ln(f_i/f_j)d\mu}$$

where the infimum is subject to the restriction that

$$(6.7) \qquad \sum_{j\neq i} \alpha_{ij} = \hat{\alpha}_{ii} \quad \text{is fixed .}$$

The reader may like to consider case (ii) on his own.

Examples. The following two examples compare the ASN of a three hypothesis test for the unknown mean of a normal distribution with two theoretical lower bounds. The test is one investigated by the author [11] and constrains the error matrix in accordance with case (ii). The lower bounds for ASN are based on (6.4) and Remark 2, respectively. θ is the true mean and $\sigma^2 = 1$ is the variance. Hypothesis H_i is that $\theta = \theta_i$, $i=1,2,3$.

Example 1:

$$\theta_1 = -.1,\ \theta_2 = 0,\ \theta_3 = .1$$

$$\alpha_{11} = \alpha_{22} = \alpha_{33} = .95\ ,\ \alpha_{21} = 1/60\ ,\ \alpha_{23} = 2/60\ .$$

θ	-.2	-.1	-.05	0	.05	.1	.2
ASN for author's test	269.5	741.2	1167	803.3	972.7	609.8	223.3
First lower bound for ASN	96.3	738.4	852.0	572.4	738.2	606.9	81.2
Second lower bound for ASN	109.3	353.8	940.0	353.6	867.5	318.5	99.3

Example 2:

$$\theta_1 = -.1,\ \theta_2 = 0,\ \theta_3 = .2$$

$$\alpha_{11} = \alpha_{22} = \alpha_{33} = .95\ ,\ \alpha_{21} = \alpha_{23} = .025$$

θ	-.2	-.1	-.05	0	.1	.2	.3
ASN for Author's test	242.5	661.8	1072	574.4	287.9	168.5	90.9
First lower bound for ASN	87.6	661.4	787.3	561.0	196.8	165.4	48.2
Second lower bound for ASN	104.0	335.0	883.3	335.0	220.8	83.8	43.0

The first lower bound is extremely good when H_1 or H_3 is true, differing from the test's ASN with errors ranging between .06% and 2%. The second lower bound does better when the true value of θ is far away from the hypothesis values. It should be remembered that the first and second lower bounds are generalizations of Wald's bound (1.1) and Hoeffding's bound (1.5). These two bounds illustrate similar behavior. When H_1 or H_2 is true, the first lower bound does somewhat better in example 2 than in example 1. This is probably due to the fact that in the second example the test is primarily a contest between H_1 and H_2 except for relatively large θ. For smaller θ, the author's test is approximately an SPRT (between H_1 and H_2) and it is well known that Wald's lower bound for ASN is very close to the ASN of the SPRT when either hypothesis is true. It does not seem likely that the true value of the second lower bound will be fully assessed until more examples of three hypothesis tests are developed.

Computing the lower bounds for the examples:

Except when one of the hypotheses is true, the first lower bound is computed from formula (6.4). The computations involve finding the infimum of a continuous convex function in two variables. Some care has to be taken because the convex function is not analytic everywhere. When one of the hypotheses is true the computations are easier.

As noted, the second lower bound for ASN follows from Remark 2 above. Since normal densities with common variance belong to the same exponential family, the modified version of theorem 5.1 applies and it follows that $B(\{1,3\}) \leq B(\{1,2,3\})$ for any error matrix A. Then, it follows that the infimum of $\max_{D\in\mathcal{D}} B(D)$, taken over the appropriate set of error matrices, is achieved when

$$A = \begin{pmatrix} \alpha_{11} & \max(\hat{\alpha}_{11}-\alpha_{23},0) & \min(\hat{\alpha}_{11},\alpha_{23}) \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \min(\hat{\alpha}_{33},\alpha_{21}) & \max(\hat{\alpha}_{33}-\alpha_{21},0) & \alpha_{33} \end{pmatrix} .$$

The lower bound is found using A with the modified version of Theorem 5.1. The computations of $R(D)$ and $T(D)$ are based on (5.28) and (5.29), respectively.

7. Lower bounds for ASN under the correct decision approach.

This section is primarily based on sections 3 and 5.

As indicated in the introduction, the correct decision approach is an alternative to the error matrix approach for choosing one among K densities. In the tradition of hypothesis testing, the parameter space Ω is partitioned into K disjoint sets $S_1,\ldots,S_K$ corresponding to K hypotheses $H_1,\ldots,H_K$ where H_i is the hypothesis that S_i contains θ. It is frequently impossible for a test to accept the correct hypothesis with high probability for all values of θ.

One is usually willing to establish "indifference regions" in the vacinity of the boundaries which say in effect that for certain θ more than one hypothesis is acceptable. Let $S_i' \supset S_i$ be the set of $\theta\epsilon\Omega$ for which a choice of H_i is acceptable, for $i=1,\ldots,K$. The acceptance of H_i is said to be a correct decision (C.D.) if S_i' contains θ.

Finally, a function $P^*(\theta)$ is specified with the requirement that a correct decision must be made with probability greater than or equal to $P^*(\theta)$ when the true parameter is θ, for each $\theta\epsilon\Omega$. It seems appropriate to refer to this requirement as the P*-condition and to the sets $S_1',\ldots,S_K'$ as the correct decision sets.

Analagous results to Theorem 3.1.

Now, suppose that f_θ is the density function under θ and that all of the density functions are with respect to the same measure μ. We will let P_θ and E_θ denote the corresponding probability measure and expectation operator under θ. Define

$$(7.1)\qquad P_{\theta_1\theta_2} \equiv P_{\theta_1}[\text{making a C.D. for } \theta_2]$$

and

$$(7.2)\qquad I(\theta) \equiv \{\text{indices } i \,|\, \theta\epsilon S_i'\} .$$

If the value of $P_{\theta_1\theta_2}$ was fixed and known for all θ_1 and θ_2 we could use the lower bound

$$(7.3) \qquad E_{\theta_o}(N) \geq \sup_{\substack{\theta \neq \theta_o \\ \theta \in \Omega}} \frac{P_{\theta_o \theta'} \ln(P_{\theta_o \theta'}/P_{\theta\theta'}) + \hat{P}_{\theta_o\theta'} \ln(\hat{P}_{\theta_o\theta'}/\hat{P}_{\theta\theta'})}{\int f_{\theta_o} \ln(f_{\theta_o}/f_\theta) d\mu}$$

for arbitrary $\theta_o, \theta' \epsilon \Omega$. (7.3) is analogous to the bound in Theorem 3.1 and can be derived in a similar manner.

Nevertheless, we do know that

$$(7.4) \qquad P_{\theta_1\theta_2} \geq P_{\theta_1\theta_1} \geq P^*(\theta_1) \quad \text{for} \quad I(\theta_1) \subset I(\theta_2) ,$$

and

$$(7.5) \qquad P_{\theta_1\theta_2} \leq \hat{P}_{\theta_1\theta_1} \leq \hat{P}^*(\theta_1) \quad \text{for} \quad I(\theta_1) \cap I(\theta_2) = \phi ,$$

where ϕ denotes the null set. Setting $\theta' = \theta$ in (7.3) and using (7.4) and (7.5), we can derive the bound

$$(7.6) \qquad E_{\theta_o}(N) \geq \sup_{\substack{I(\theta) \cap I(\theta_o) = \phi \\ P^*(\theta) + P^*(\theta_o) \geq 1 \\ \theta \epsilon \Omega}} \frac{\hat{P}^*(\theta_o) \ln(\hat{P}^*(\theta_o)/P^*(\theta)) + P^*(\theta_o) \ln(P^*(\theta_o)/\hat{P}^*(\theta))}{\int f_{\theta_o} \ln(f_{\theta_o}/f_\theta) d\mu}$$

This may be derived by observing that

$$\{\theta\epsilon\Omega: \theta \neq \theta_o\} \supset \{\theta\epsilon\Omega: I(\theta) \cap I(\theta_o) = \phi , P^*(\theta) + P^*(\theta_o) \geq 1\} ,$$

by taking the infimum of the right-hand side of (7.3) subject to the constraints imposed by (7.4) and (7.5), then interchanging "inf" and "sup", and finally, by observing the monotonicity of the function

$x\ln(x/y) + \hat{x}\ln(\hat{x}/\hat{y})$ when $x \leq y$. A more careful analysis yields the slightly better lower bound

$$(7.7) \quad E_{\theta_o}(N) \geq \inf_{\{b_j : B(b,\theta_o)=P^*(\theta_o)\}} \sup_{\substack{I(\theta)\cap I(\theta_o)=\emptyset \\ P^*(\theta)+P^*(\theta_o)\geq 1 \\ \theta\in\Omega}} \frac{B(b,\theta)\ln(B(b,\theta)/P^*(\theta))+\hat{B}(b,\theta)\ln(\hat{B}(b,\theta)/\hat{P}^*(\theta))}{\int f_{\theta_o}\ln(f_{\theta_o}/f_\theta)d\mu} ,$$

where $b_j \equiv P_{\theta_o}$ (accepting H_j), $j=1,\dots,K$, and where $B(b,\theta') \equiv \sum_{i\in I(\theta')} b_i = P_{\theta_o\theta'}$ for arbitrary $\theta'\in\Omega$.

Analogous results to theorem 5.1.

Suppose that the density function f_θ is of the exponential form $c(\theta)h(x)e^{Q(\theta)t(x)}$ and that $Q(\theta)$ is strictly monotone in real valued θ. Then f_θ is pairwise minimizable and, in fact,

$$\min_{\substack{a\leq\theta\leq b \\ \theta\in\Omega}} f_\theta(x) = \min(f_a(x),f_b(x)) \quad \text{for} \quad a,b\in\Omega .$$

It becomes appropriate to redefine R, S and T (used in Theorem 5.1) for intervals [a,b] instead of index sets D.

(7.8) $$R[a,b] \equiv \sup_{\substack{a \leq \theta \leq b \\ \theta\in\Omega}} \int f_{\theta_o} \ln(f_{\theta_o}/f_\theta)d\mu$$

$$= \max(\int f_{\theta_o} \ln(f_{\theta_o}/f_a)d\mu,\ \int f_{\theta_o} \ln(f_{\theta_o}/f_b)d\mu) .$$

The latter equality holds because $-\ln f_\theta$ is convex in $Q(\theta)$.

(7.9) $$T^2[a,b] \equiv \frac{1}{16} \int f_{\theta_o} [\ln(f_b/f_a)-E_{\theta_o}(\ln(f_b/f_a))]^2 d\mu,$$

making

(7.10) $$T[a,b] = \frac{1}{4}|Q(b)-Q(a)| \cdot \mathrm{Var}_{\theta_o}(t(X)) .$$

(7.11) $$S[a,b] \equiv \sum_{j=1}^{K} \inf_{\substack{a \leq \theta \leq b \\ \theta\in\Omega}} \alpha_{\theta j} ,$$

where $\alpha_{\theta j} = P_\theta$ [accepting H_j] for $j=1,\ldots,K$.

If $\alpha_{\theta j}$ were fixed and known for $\theta\in\Omega$ and $i=1,\ldots,K$, then we would have the lower bound

(7.12) $$E_{\theta_o}(N) \geq [(T^2[a,b]-R[a,b]\ln S[a,b])^{1/2}-T[a,b]]^2/R^2[a,b] .$$

Since the fixing of $\alpha_{\theta j}$ constitutes more control of the errors than is implied in the P*-condition, we must take an infimum over the class of all sets $\{\alpha_{\theta j}\}$ which satisfy that condition. We shall not treat this problem any further than to note that the values of R and T are independent of the errors and to note also that the problem is primarily one of finding the supremum of $S[a,b]$ taken over the same class of error sets. Calling this supremum $S^*[a,b]$, we get the correct lower bound

$$(7.13) \qquad E_{\theta_0}(N) \geq [(T^2[a,b]-R[a,b]\ln S^*[a,b])^{1/2}-T[a,b]]^2/R^2[a,b] \ .$$

Actually, one might legitimately take a supremum of the right-hand side of (7.12) over intervals $[a,b]$ before taking an infimum over sets $\{\alpha_{\theta j}\}$, but this makes computations more difficult.

Example. This example is based on an example used by Sobel and Wald [12] in their paper concerning a three hypothesis sequential test for the unknown mean of the normal distribution. They require a set of constants $-\infty < \theta_1 < a_1 < \theta_2 \leq \theta_3 < a_2 < \theta_4 < \infty$ to define the two sets S_1, S_2, S_3 and S_1', S_2', S_3'. Then $S_1 = (-\infty, a_1)$, $S_2 = [a_1, a_2]$, $S_3 = (a_3, \infty)$, and $S_1' = (-\infty, \theta_2)$, $S_2' = (\theta_1, \theta_4)$, $S_3' = (\theta_3, \infty)$.

$I(\theta)$ and $B(b,\theta)$ are given as follows:

range of θ	$I(\theta)$	$B(b,\theta)$
$-\infty < \theta \leq \theta_1$	$\{1\}$	b_1
$\theta_1 < \theta < \theta_2$	$\{1,2\}$	$b_1 + b_2$
$\theta_2 \leq \theta \leq \theta_3$	$\{2\}$	b_2
$\theta_3 < \theta < \theta_4$	$\{2,3\}$	$b_2 + b_3$
$\theta_4 \leq \theta < \infty$	$\{3\}$	b_3

In their problem, $P^*(\theta) = c$, a constant greater than .5, and $\int f_{\theta_o} \ln(f_{\theta_o}/f_\theta)d\mu = (\theta_o-\theta)^2/2\sigma^2$, where $\sigma^2 = 1$ is the common variance. In terms of the function

$$h(x,y) = x \ln(x/y) + \hat{x} \ln(\hat{x}/\hat{y}) \quad ,$$

we can express lower bounds (7.6) and (7.7) as follows:

range of θ_o	lower bound (7.6)	lower bound (7.7)
$-\infty < \theta_o \leq \theta_1$	$2(\theta_o-\theta_2)^{-2}\, h(\hat{c},c)$	the same
$\theta_1 < \theta_o < \theta_2$	$2(\theta_o-\theta_4)^{-2}\, h(\hat{c},c)$	the same
$\theta_2 \leq \theta_o \leq \theta_3$	$2\max[(\theta_o-\theta_1)^{-2},(\theta_o-\theta_4)^{-2}]h(\hat{c},c)$	$2 \inf\limits_{0<b_1<\hat{c}} \max[\frac{h(b_1,c)}{(\theta_o-\theta_1)^2}, \frac{h(\hat{c}-b_1,c)}{(\theta_o-\theta_4)^2}]$
$\theta_3 < \theta_o < \theta_4$	$2(\theta_o-\theta_1)^{-2}\, h(\hat{c},c)$	the same
$\theta_4 < \theta < \infty$	$2(\theta_o-\theta_3)^{-2}h(\hat{c},c)$	the same

Sobel and Wald considered the following case [1]:

$$\theta_4 = -\theta_1 = 5/16 \ , \ \theta_3 = -\theta_2 = 3/16 \ \text{ and } \ c = .97101 \ .$$

They did not derive an exact formula for their ASN but derived an upper and lower bound for it. It should be emphasized that their upper and lower bounds are not universal but apply to their test only.

1/ For certain reasons of accuracy the author chose to use c = .97101 instead of their value of c = .971. Since the values in the table are quite sensitive to the value of c, there is a small but apparent discrepancy between the two tables.

θ	0	1/16	2/16	3/16	4/16	5/16	6/16	8/16	10/16
Their upper bound	146.2	163.6	229.6	425.3	790.9	425.1	224.7	112.4	74.9
Their lower bound	112.4	149.8	224.3	423.4	789.2	423.4	224.3	112.4	74.9
lower bound (7.6)	67.7	105.9	188.2	423.4	20.9	423.4	188.2	67.7	34.6
lower bound (7.7)	69.9	105.9	188.2	423.4	20.9	423.4	188.2	67.7	34.6
lower bound (7.13)	49.3	73.9	122.6	225.9	645.2	240.5	122.6	49.3	26.4

Lower bound (7.13) was computed using the interval $[\theta_3, \theta_4]$. It may be checked that $S^*[\theta_3, \theta_4] = \frac{3}{2}(1-c)$.

REFERENCES

[1] Anderson, T.W. (1960). "A modification of sequential analysis to reduce the sample size." Ann. Math. Statist., **31**, 165-197.

[2] Blackwell, D. (1946). "On an equation of Wald." Ann. Math. Statist., **17**, 84-87.

[3] Blackwell, D. and Girshick, M.A. (1946). "On functions of sequences of independent chance vectors with applications to the problem of random 'walk' in K dimensions." Ann. Math. Statist., **17**, 310-317.

[4] Bohnenblust, H.F., Karlin, S. and Shapley, L.S. (1950). "Games with continuous, convex pay-off." Contributions to the Theory of Games, **1**, 181-192. Princeton University Press.

[5] Chow, Y.S., Robbins, H., and Teicher, H. (1965). "Moments of randomly stopped sums." Ann. Math. Statist., **36**, 789-799.

[6] Doob, J.L. (1953). Stochastic Processes. Wiley.

[7] Hoeffding, W. (1953). "A lower bound for the average sample number of a sequential test." Ann. Math. Statist., **24**, 127-130.

[8] Hoeffding, W. (1960). "Lower bounds for the expected sample size and the average risk of a sequential procedure." Ann. Math. Statist., **31**, 352-368.

[9] Paulson, E. (1963). "A sequential decision procedure for choosing one of K hypotheses concerning the unknown mean of a normal population." Ann. Math. Statist., **34**, 549-554.

REFERENCES

[10] Rao, C.R. (1952). Advanced Statistical Methods of Biometric Research. Wiley,

[11] Simons, G.D. (1966). "Multihypothesis testing." Ph.D. Thesis, Department of Statistics, University of Minnesota.

[12] Sobel, M. and Wald, A. (1949). "A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution." Ann. Math. Statist., **20**, 502-522.

[13] Wald, A. (1944). "On cumulative sums of random variables." Ann. Math. Statist., **15**, 283-296.

[14] Wald, A. (1945). "Sequential tests of statistical hypotheses. Ann. Math. Statist., **16**, 117-186.

[15] Wald, A. (1946). "Some improvements in setting limits for the expected number of observations required by a sequential probability ratio test." Ann. Math. Statist., **17**, 466-474.

[16] Wald, A. (1946). "Differentiating under the expectation sign in the fundamental identity of sequential analysis." Ann. Math. Statist., **17**, 493-497.

[17] Wald, A. and Wolfowitz, J. (1948). "Optimum character of the sequential probability ratio test." Ann. Math. Statist., **19**, 326-339.

[18] Wolfowitz, J. (1947). "The efficiency of sequential estimates and Wald's equation for sequential processes." Ann. Math. Statist., **18**, 215-230.